Discerning excellence from mediocrity in swimming: new insights using Bayesian quantile regression.

Tony D. Myers[1,] Yassine Negra[2], Senda Sammoud[2], Helmi Chaabene[4], Alan M. Nevill[3],

[1] *Sport and Health, Newman University, Birmingham, UK*

[2] *Research Unit (UR17JS01) «Sports Performance, Health & Society», Higher Institute of Sport and Physical Education of Ksar Saîd, Universite de la Manouba, Tunisia.*

[3] *Faculty of Education, Health and Wellbeing, University of Wolverhampton, UK*

[4] *Division of Training and Movement Sciences, Research Focus Cognitive Sciences, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany.*

Abstract

Purpose: Previous research has captured point estimates for population means of somatic variables associated with swimming speed across strokes, but have not determined if predictors of swimming speed operate the same at the upper tails of the distribution (τ =0.9) as they do at the median levels (τ =0.5) and lower levels (τ =0.1). **Method:** Three hundred sixty-three competitive-level swimmers (male [n=202]; female [n=161]) participated in the study. To identify key somatic variables associated with 100-m swimming across and between strokes controlling for age, we used a Bayesian allometric quantile regression model, refined using Bayes Factors and Leave-one-out cross validation. **Results:** High probabilities (>99%) were found for arm-span, seated-height and shoulder-breadth being the strongest somatic predictors across strokes. For individual strokes, Bayesian quantile regression demonstrated that the relative importance of predictors differs across quantiles. For swimmers in the 0.9 quartile, shoulder-breadth is a more important than height for front-crawl, wide shoulders are important for breaststroke swimmers but can be detrimental when combined with narrow hips, seated-height and hip-width are

important for backstroke swimming speed, and calf girth for butterfly. **Conclusion:** These results highlight the importance of considering key somatic variables for talent identification in swimming and ensure young swimmers focus on strokes compatible with their somatic structure. The most important new insight is that predictors differ for the best swimmers compared to average or poorer swimmers. This has implications beyond swimming, pointing to the importance of considering the upper tails of distributions in performance and talent identification contexts.
Keywords: swimming; somatic variables; talent; Bayesian modelling; allometric modelling

1. Introduction

Identifying key somatic characteristics of athletes from an early age is an important tool for talent identification.[1,2]. This is particularly true in swimming, where they are considered amongst the most important factors in enabling swimmers to achieve a high-level performance during their careers [3,4]. While for talent identification purposes, there are concerns with how limb sizes may change with maturation, proportional body segment breadth measures are stable from preadolescence through to adulthood, and the proportionality of upper limbs stable from mid-adolescence onwards [5]. This stability in proportion seems to point to the importance of using allometric models when exploring somatic variables in the context of talent identification. While standard linear analysis can fail to properly represent proportional changes, allometric models where both response Y and predictors X are log transformed, are able to explain proportionality between the Y and X. In a performance context, these models also have the advantage of having the flexibility of a non-linear quadratic in age within an exponential term ensuring that speed remain non-negative irrespective of the child or adolescent's age.

A number of studies have explored the morphology of swimming, examining somatic and demographic predictors of particular swimming strokes [6,7,8]. A swimmer's morphology is genetically determined, with inheritance for body and limb length around 70%, and 50% for body and limb girths respectively [9]. Therefore, it seems sensible, where possible, to encourage young athletes to make the most of their inherited morphology by guiding them to train and compete in strokes that are the most compatible with their particular somatic structure.

The primary method of analysis in these studies is ordinary least squares regression. While this provides a good estimate of y when x is close to the mean, it can perform poorly at other levels [10]. Nonetheless, exploring the key predictors for the very best athletes is really important in elite sport and talent development contexts, therefore alternative analysis methods that can identify predictors of the best rather than average performers in any particular cohort may prove more useful. A recent study attempted to explore differences in some somatic variables according to swimming speed [11]. Swimmers were grouped into one of three levels: level 1, national champions, national record holders and/or enrolled in a talent ID program; level 2, swimmers racing in national competitions; and level 3, swimmers racing predominantly in local and regional competitions. The study found level 1 and level 2 athletes were, on average, taller, heavier and had longer limbs than those in level 3.

No previously published study has explored if predictors of swimming performance operate the same at the upper tails of the distribution as they do at the mean or median levels. Quantile regression allows this by quantifying the relationship of important predictors across different aspects of the conditional distribution of swimming speed, while also allowing error heteroscedasticity [12]. Therefore, this study aimed to determine the key somatic variables associated with the best young swimmers in each of the four swimming strokes using this approach, to allow coaches to guide young swimmers to train and compete in strokes that are the most compatible with their somatic structure. The best young swimmers are defined as those in the 90th percentile of the cohort for swimming speed and poorest in the 10th percentile, after controlling for age in each case.

2. Methods

2.1 Participants

Three-hundred and sixty-three competitive-level swimmers (male [n=202]; female [n=161]) were recruited for the study (front-crawl swimmers: n=74, butterfly swimmers: n=167, backstroke swimmers: n=63, and breaststroke swimmers: n=59), mean age =13.85± 3.10. Mean age for the individual strokes were, front crawl = 17.14 ± 3.53, butterfly = 13.29 ± 2.75, backstroke =13.48 ± 1.07, and breaststroke = 11.69 ± 1.20 (all demographic details are outlined in supplementary Table 1). We acknowledge that these data have been published previously [13], but critically, the focus, analysis and findings for that study were markedly different. The reason for differences in the mean age of participants across strokes is because of national requirements for swimmers under 13 years of age to participate in butterfly, backstroke and breaststroke, but not front crawl.

Swimmers were recruited from 4 clubs in the capital of Tunisia one week prior to the winter championship. They were all competing at a national level and following the same volume of training for each age category (the volume required by the Tunisian Swimming Federation). In addition, a number of the participants competed at international meetings including the Word-Cup, and 13 participants were medallists (gold, silver and bronze) in Maghrebin (North African) championship.

At the time of data collection, participants were swimming training five to six training times per week (4000 ± 1000 m per session; 8 ± 1 hour per-week) in preparation for the Tunisian Winter National Championships. The total distance per session for each stroke was 5000 ±1000 m for the front crawl, and 4000 ±1000 m, for the breaststroke, backstroke, and butterfly. In addition, to the swimming training, swimmers were involved in 3 dry-land training sessions per week (4±1 hour per week).

Written informed parental consent and participant consent or assent was obtained prior to the start of the study. All youth athletes and their parents / legal guardians were informed about the experimental protocol and its potential risks and benefits before the commencement of the research. Institutional ethical approval was gained from the Ethics Institutional Review Committee for the ethical use of human subjects at Ksar Saïd University, Tunisia.

2.2 Somatic measurements

All the somatic measurements were taken in accordance with standardized procedures of the international society for the advancement of kinanthropometry (ISAK)[14] (Supplementary Table 1).

After calibration, testing was undertaken in a standardized order. Height (m) and body-mass (kg) were assessed to the nearest 0.1 cm and 0.1 kg, using a SECA stadiometer and a SECA weighing scale (SECA Instruments Ltd, Hamburg, Germany). Skinfold measurements (in millimetres) were made using Harpenden skinfold callipers (Harpenden Instruments, Cambridge, UK). Skinfold measurements were taken on the right-hand side of the body at the triceps and the subscapular, with body-fat estimated using the Slaughter et al.[15] skinfold equation. Limb-lengths, girths and breadths were measured using a large sliding calliper and anthropometric measuring tape. Each assessed via direct measures using landmarks techniques.

Upper arm length was measured from landmarks placed to acromiale and dactylion while athletes stood in the erect position. The lower-arm length was measured as the distance between measured from radiale and stylion landmarks. Hand-length was measured as the shortest distance from the marked midstylion line to the dactylion. Upper-limb length was taken as the distance between marked acromiale and radiale landmarks. Lower limb length was determined by subtracting sitting height from standing height. Thigh-length was determined as the distance between trochanterion and tibiale lateral landmarks. Leg length was the distance from the height of the tibiale lateral to the top of the surface where the participant stood. Foot-length was measured from the Akropodion (i.e., the tip of the longest toe which may be the first or second phalanx) to the Pternion (i.e., posterior point on the calcaneus of the foot). Arm-relaxed girth was measured at the marked level of the mid-acromiale- radiale. The tape was positioned perpendicular to the long axis of the arm.

Forearm-girth was measured at the maximum girth of the forearm distal to the humeral epicondyles. Wrist-girth was taken distal to the styloid processe; the minimum girth in this region. Thigh-girth was taken at the marked mid-trochanterion-tibiale-lateral site. Calf-girth was the maximum girth of the calf taken at the marked medial calf skinfold site. Ankle girth was determined as the minimum girth of the ankle taken at the narrowest point superior to the Sphyrion tibiale. Biacromial breadth was measured between the most lateral points of the acromion processes. Bi-iliocristal breath was defined as the distance between the most lateral points on the iliac crests. All somatic measures were recorded twice and the mean scores were retained for the statistical analysis. Intraclass correlation coefficients (ICCs) for test-retest reliability ranged from 0.97 to 0.99 for all somatic and skinfolds measures.

2.3 Swimming speed quantification

The swimming speed (SS), in meters per second ($m.s^{-1}$), was calculated from swimming times in seconds from the official results published by the Tunisian swimming Federation during the Winter National Championships. The average speed was calculated as the ratio between distances swam and the total time recorded in this distance ($m.s^{-1}$). at the championships, SS was measured with a high technology electronic timing (Omega, Switzerland). Water temperature, as determined by Fédération Internationale De Natation [16], was kept between 25 and 28 degrees centigrade.

Descriptive statistics (means ± SD) of SS, demographic and somatic measurements by sex and stroke are presented in supplementary Table 1.

2.4 Data analysis

Allometric modelling was used for all models fitted, which involves log-transforming response and predictor variables. This log-transformation is useful for representing proportional changes in allometric relationships, something that is essential for the present study. A Bayesian approach was taken to data analysis, in part to circumnavigate the widely reported issues with the misinterpretation of traditional p-values [17], but also because of the more intuitive interpretation of the results in terms of direct probability and statistical intervals. Moreover, Bayesian models provide an estimate of a complete distribution rather a single value and this allows more nuanced consideration of parameters.

The data analysis was conducted in two stages. The first stage involved identifying key predictors across all swimming strokes combined, and the second involved identifying key predictors of each stroke separately. For the first stage of analysis, a saturated Bayesian allometric regression model was fitted with all predictors included. The measurements, used as predictors in the model, included body-mass, height, percentage body-fat and limb dimensions (lengths and girths). In order to determine the best predictors of swimming speed, this initial model was fitted using a Jeffrey's prior on sigma and a Zellner-Siow Cauchy prior on model coefficients. The aim of this being to select the combination of predictors with the highest Bayes Factor (BF). Bayes Factors can be used to identify models with the highest amount of evidence in their favor from the models considered [18]. Marginal posterior inclusion probabilities (MPIP) were calculated to determine how likely a particular predictor was in the 'true model'. Bayes Factors, and MPIP for models and variables were calculated using the Bayesian adaptive sampling algorithm described by Merlise, Ghosh, and Littman [19] and implemented using the Bayesian Adaptive Sampling

(BAS) package [20] in R [21]. Posterior inclusion probabilities greater than 0.5 were included in the model and any predictor with a lower probability was discounted.

The second stage of analysis involved modeling the best predictors for each stroke in three phases. Phase one, involved identifying the best predictors using the same methods described above. In phase two, the predictors identified in phase one were further refined by fitting different response distributions to each model, including Gaussian and t-distributions. To determine which of these models best fitted the data, Leave-One-Out cross-validation (LOO) was used. LOO uses log-likelihoods from posterior simulations of the parameter values to estimate point-wise out-of-sample prediction accuracy to determine the relative predictive performance of the model to the data; the lowest LOO information criterion (LOOIC) the higher the predictive accuracy [22]. Once the best models for each stroke – in terms of predictors and response distributions - had been identified, phase three involved fitting Bayesian quantile regression models for each stroke (0.1, 0.5 and 0.9 quantiles) to understand how the key predictors explain swimming speed (SS) in the fastest swimmers compared to slower swimmers at the middle and bottom of the distribution. All models were fitted using the Bayesian Regression Models using Stan (brms) package [23] with MCMC sampling via Stan [24].

Along with posterior distributions for parameters in model, the probability of direction for each parameter in each model was calculated. Mathematically, the probability of direction is defined as the proportion of the posterior distribution that has the same sign as the median, and can be interpreted as the probability- expressed as a percentage- that a parameter described by its posterior distribution is strictly positive or negative (whichever is the most probable). Moreover, given it familiarity in standard regression analysis, a Bayesian version of R^2 was also calculated as an estimate of the proportion of variance explained for future performances [25].

All models reported were checked for convergence ($\hat{r} = 1$), with the graphical posterior predictive checks showing simulated data under the best fitted models compared well to the observed data with no systematic discrepancies. To illustrate how the relationships between different somatic variables and SS in the four different strokes, the 0.9 quantile models were used to make predictions for SS and 100-m times using new but plausible values for predictors within the range of the empirical data collected.

3. Results

3.1 Across strokes

Predictors with marginal inclusion probabilities greater than 0.5 included log transformed seated height (0.60), log transformed bi-acromial breadth (0.99), log transformed bi-iliac breadth (0.77), log transformed arm span (0.98), log transformed body fat (1) sex (0.54), and age (0.97).

The strongest somatic predictors across strokes are shoulder width (bi-acromial breadth) followed by arm-span and seated-height. For every one percent increase in shoulder-width we can expect SS to increase by 0.42 percent, 0.33 percent for arm-span, and 0.30 for seated-height. There are extremely high probabilities of positive relationships between 100-m SS and shoulder-width (100%) arm span (100%), and seated-height (99.95%) conditional on the data (see Table 1).

****Table 1 near here****
****Table 2 near here****

3.1.1 Front-crawl

The regression slopes are different for different quantiles (0.1, 0.5, and 0.9) of front-crawl SS (see Table 2; Supplementary figure A). For the average swimmer (τ =0.5), height is the most important variable in predicting 100-m front-crawl SS with every one percent increase in height predicting a 0.67 percent increase in SS. However, in the best swimmers (τ =0.9), shoulder-breadth (bi-acromial breadth) plays a greater role (see Table 2). For average swimmers, every one percent increase in shoulder breadth we can expect SS to increase by 0.55 percent, whereas for the faster swimmers (τ =0.9) a one percent increase results in a 0.77 percent increase in 100-m SS. The percentage of body fat is a stronger predictor for better swimmers than for average or poorer swimmers. The probability of direction for all somatic predictors for front-crawl is >96%. How different combinations of height and shoulder-breadth (bi-acromial breadth) predict 100-m SS and swimming times are shown in Table 3.

3.1.2 Breaststroke

Again, the regression slopes are different for different quantiles (0.1, 0.5 and 0.9) of breaststroke SS (see Table 2; Supplementary Figure C), albeit only minimally for shoulder-breadth (bi-acromial breadth). The influence of hip-width (bi-iliac breadth) in particular, differs across levels of SS. For the poorer swimmers (τ =0.1) having wider hips is the most important predictor of 100-m breaststroke SS, in contrast for the best swimmers (τ =0.9), shoulder-breadth is the strongest predictor. Percent body fat is a slightly stronger predictor for average breaststroke swimmers than for the best or poorest swimmers (see Table 2). The probability of direction for all somatic predictors of breaststroke in the model is >99%. The importance of different shoulder and hip breadths on100-m backstroke SS and times is illustrated by selected predictions shown in Table 3.

3.1.3 Backstroke

In backstroke SS, once again, the regression slopes differ across the different quantiles (0.1, 0.5 and 0.9) (see Table 2; Supplementary figure B). Arm-span is the most important somatic variable in predicting backstroke SS in poor (τ =0.1) and average (τ =0.5) swimmers, but least important for the best swimmers (τ = 0.9), where seated-height is the strongest predictor. In the strongest backstroke swimmers, for a one percent increase in their seated-height, we can expect a 0.49 percent increase in 100-m SS. Percent Body-fat has a slightly weaker relationship with 100-m breaststroke SS in the weaker and average swimmer than with the best swimmers, but only minimally so. The probability of direction of the majority of somatic predictors of backstroke SS were <99%, with the exception of arm-span (90.85%). The importance of seated-height and hip-width on 100-m backstroke SS and times are illustrated by the predictions in Table 3.

3.1.4 Butterfly

As with all other strokes, the regression slopes differ across the different quantiles (0.1, 0.5 and 0.9) of 100-m butterfly SS (see Table 2; Supplementary figure D). For lower performing swimmers (τ =0.1) shoulder-breadth (bi-acromial breadth) is the strongest somatic predictor of 100-m butterfly SS, with a one percent increase in shoulder-breadth 100-m butterfly SS is predicted to increase by 0.62 percent. However, shoulder-breadth is the weakest of the somatic predictor for the best swimmers (τ =0.9). For both the average (τ =0.5) and the best swimmers (τ =0.9), calf girth is the strongest single predictor. For the average swimmer, a one percent increase is associated with a 0.82 percent increase in 100-m butterfly SS., and with the best swimmers a 0.45 percent increase in SS. Interestingly, the difference between ankle girth and calf girth is less pounced in the best swimmers compared to those in the lower quartiles (τ =0.1 and 0.5). Percent body fat is a predictor across quantiles, but is a slightly less important predictor for the best swimmers. The probability of direction of the majority of the somatic predictors are >95%, with the exception of bi-acromial breadth (90.77%). The impact of different combinations of calf girth, ankle girth and hip-width (bi-iliac breadth) measurements in predicting 100-m butterfly SS and time are shown in Table 3.

****Table 3****

4. Discussion

This study aimed to determine the key somatic variables associated with the best young swimmers in each of the four swimming strokes, to allow coaches or other interested others to guide young swimmers to train and compete in strokes that are the most compatible with their particular somatic structure. The most important and novel finding of the present study is that somatic predictors of SS differ in the upper tails of the distribution in each of the four strokes, suggesting the relative importance of predictors differ for the best swimmers compared to average or poorer swimmers. This has implications beyond swimming, pointing to the importance of considering the upper tails of distributions in performance contexts in general and for talent identification specifically.

Bayesian allometric modeling was initially used to identify the optimal somatic measurements associated with 100-m SS across swimming strokes. Then, importantly for the purpose of talent identification, identifying the most important somatic predictors for the best swimmers in each stroke using Bayesian quantile regression.

After controlling for age and sex, five key somatic predictors were identified as having a high probability (>96%) of an association with average SS (m.s-1) across the four strokes. The strongest association was with shoulder breadth, followed by arm-span, seated-height, body fatness and least convincingly hip breadth. This suggests, it is an advantage in general for swimmers to have broader shoulders, a longer torso, long arms, low body fatness and, far less certainly, narrower hips. These general characteristics have been identified previously [26, 27, 1]. It is a generally acknowledged that arms contribute more to force generation than the legs in swimming [28, 29, 30, 31], with broad shoulders and a long arm reach useful for a swimmer across strokes, especially when combined with a long torso. A long torso, particularly relative to leg length, shifts a swimmer's centre of mass toward their hips, achieving a more horizontal body position improving hydrodynamic efficiency by reducing drag and allowing for maximum propulsion, the centre of mass and the centre of buoyancy being closely related [32]. Similar to other sports, our findings suggest lower body fatness does play a role in SS across strokes.

4.1 Front crawl

The results clearly show how height and shoulder breath are important considerations for practitioners to consider when selecting swimmers for front crawl training, something that has been identified previously [33]. Nonetheless, the new insight from quantile regression is that for the best swimmers' shoulder breadth is a more important consideration than height per se. The predictions from the front crawl model clearly illustrate this (see Table 3). For example, all else being equal, a front crawl swimmer who stands 195-cm tall, would take 0.51s less time to swim 100-m than a swimmer 5-cm shorter. However, importantly, a swimmer with the same height and only 1-cm lower arm span, is predicted to swim 100-m 0.65 sec slower. So, a 1-cm difference in shoulder breadth has a greater impact on SS than a 5-cm difference at this height (see Table 3).

4.2 Breaststroke

The somatic predictor with the strongest relationship with SS in the poorest swimmers is hip-width followed by shoulder-breadth. While shoulder width has a stronger relationship to SS than hip-width in the best swimmers, the more interesting finding concerns the relationship between these two measures. While not obvious from the regression coefficients, the predictions show a pronounced V-shape - wider shoulders and narrow hips - is related to poorer breaststroke SS in the best breaststroke swimmers (see Table 3).

4.3 Backstroke

The somatic predictor with the strongest relationship with SS in the best backstroke swimmers is seated-height, followed by hip-width. Nonetheless, assuming importance directly from this can be misleading. Allometric scaling is relative to the attribute being measured. In the sample, the seated-height of backstroke swimmers covers 24.3-cm, whereas, hip breadth ranges are much less at 9.5-cm between lowest to highest. Given the effects are multiplicative, a 1-cm difference in hip breadth has more of impact than a 1-cm difference in seated-height (see Table 3).

4.4 Butterfly

Calf girth is the strongest somatic predictor for across quartiles in butterfly swimmers. This suggests for the best, average and poorest butterfly swimmer, lower leg muscularity is important in this stroke and may be related to the strong knee flexion required for and optimal propulsion and hip position [34]. However, given doubts as to the suitability of using lower limb proportions as selection criteria in talent identification programs [5], calf muscularity may not be useful for selection purposes. For slower and average swimmers, shoulder-breadth is an important predictor of SS. Nonetheless, our findings suggest that shoulder-breadth is less important for the best swimmers (see Table 3).

As expected, this study has some limitations. Firstly, our models do not include predictors such as muscle genetics (muscle fibre type etc.) or technique quality, which may mediate the relationship between the somatic variables and SS. Nonetheless, to include these measurements, participant numbers would need to be reduced which would be detrimental for the study. Future research should consider exploring the relationship between somatic variables and SS across different distances (50-m, 200-m etc.). Finally, we acknowledge that given body proportions differ in pre-pubertal and early-pubertal children when compared with adolescents, our predictions for the ideal proportions at different ages may be influenced by these differences.

5. Conclusion

A number of studies have explored the somatic and demographic predictors of SS, but none have explored how the importance of each predictor might change when comparing the upper tails of the distribution to the usual mean or median levels. The results support some the results of some previous findings in highlighting importance of somatic measurements for talent identification and/or athlete monitoring purposes. Nonetheless, crucially the findings of this study highlight the importance of considering the upper tails of the distribution when exploring predictors. In practical terms, while arm-span, seated-height and shoulder-breadth were important across the four swimming strokes investigated, for individual strokes, the relative importance of predictors differed for the best compared to the average and poorest swimmers. We found shoulder-breadth to be a more important consideration than height for front crawl; wider shoulders combined with narrow hips to be detrimental for the best breaststroke swimmers; seated-height important for backstroke SS, with hip-width also proving crucial; finally, we found calf girth important in butterfly swimmers.

In summary, our findings suggest that researchers attempting to predict sports performance in general and for talent identification in particular, the upper tails of distributions should be explored as well as the usual average or median values.

References

1. Morais, JE, Silva, AJ, Marinho, DA, Seifert, L, Barbosa, TM. Cluster stability as a new method to assess changes in performance and its determinant factors over a season in young swimmers. *Int. J. Sports Physiol. Perform*. 2015; 10, 261–268.
2. Morais JE, Silva AJ, Marinho DA, Lopes VP, Barbosa TM. Determinant factors of long-term performance development in young swimmers. *Int J Sports Physiol Perf.* 2017; 12: 198-205
3. Geladas, ND, Nassis, GP, Pavlicevic, S. Somatic and physical traits affecting sprint swimming performance in young swimmers. *Int. J. Sports Med.* 2005; 26, 139–144.
4. Lätt E, Jürimäe J, Mäestu J, Purge P, Rämson R, Haljaste K, Keskinen KL, Rodriguez FA, Jürimäe T. Physiological, biomechanical and anthropometrical predictors of sprint swimming performance in adolescent swimmers. *J Sports Sci Med.* 2010; 9:398–404.
5. Ackland, TR, Bloomfield, J. Stability of human proportions through adolescent growth. *Aust J Sci Med Sport,* 1996; 28, 2, 57-60.
6. Nevill, A. M., Oxford, S. W., & Duncan, M. J. Optimal body size and limb length ratios associated with 100-m personal-best swim speeds. *Med Sci Sports Exerc,* 2015; 47, 1714-8,
7. Sammoud, S, Nevill, AM, Negra, Y, Bouguezzi, R, Chaabene, H, Hachana, Y. 100-meter Breaststroke swimming performance in youth swimmers: the predictive value of anthropometrics. *Pediatr Exerc Sci,* 2018; 1, 30, 2017; 3393-401.
8. Sammoud S, Nevill AM, Negra Y, Bouguezzi R, Chaabene H, Hachana Y. Allometric associations between body size, shape, and 100-m butterfly speed performance. *J Sports Med Physi Fitness.* 2018; 58, 5, 630-7.
9. Issurin VB. Evidence-Based Prerequisites and Precursors of Athletic Talent: A Review. *Sport Med.* 2017: 11

10. Koenker, R. *Quantile Regression* (Econometric Society Monographs). Cambridge: Cambridge University Press. 2005.
11. Barbosa TM, Bartolomeu R, Morais JE, Costa MJ. Skillful swimming in age-groups is determined by anthropometrics, biomechanics and energetics. *Front Physiol*. 2019; 10:73.
12. Yang, YW., Wang, HJ., He, XM. Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood, *Int Stat Rev*, 2016; 84, 327–34.
13. Nevill, AM., Yassine, N, Myers, TD., Sammoud, S., & Chaabene H. Key somatic variables associated with, and differences between the 4 swimming strokes, *J Sports Sci*., 2020; 1:8
14. Stewart A, Marfell-Jones, M Olds T, de Ridder H. International Standards for Anthropometric Assessment. *Lower Hutt, New Zealand: ISAK*. 2011; 57–72.
15. Slaughter MH, Lohman TG, Boileau RA, Horswill CA, Stillman RJ, VanLoan MD, Bemben DA. Skinfold equations for estimation of body fatness in children and youth. *Hum Biol.* 1988; 60: 709-23.
16. Federation Internationale De Natation (FINA). Talent identification programmes. 2014 [cited 18 September 2014]. Available from: http://www.fina.org/.
17. Wasserstein, RL. & Lazar, NA. The ASA Statement on p-Values: Context, Process, and Purpose, *Am Stat.* 2016; 70:2: 129-133
18. Kass, R. E & Raftery, A.E. Bayes Factors, *J Am Stat Assoc*, 1995; 90:430, 773-795,
19. Merlise AC, Ghosh, J, Littman, ML, Bayesian Adaptive Sampling for Variable Selection and Model Averaging, Journal of Computational and Graphical Statistics, 2011; 20,1, 80-101.
20. Merlise, AC. BAS: Bayesian Variable Selection and Model Averaging using Bayesian Adaptive Sampling, R package version 1.5.3. 2018.
21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2019.
22. Vehtari, A., Gelman, A., & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput,* 2016; 27, 5.

23. Bürkner, P.C Advanced Bayesian Multilevel Modeling with the R Package brms. *R J,* 2018; 10, 1, 395-411.
24. Stan Development Team. RStan: the R interface to Stan. R package version 2.18.2. http://mc-stan.org/. 2018.
25. Gelman, A., Goodrich, B, Gabry, J. & Vehtari, A. R-squared for Bayesian Regression Models, *Am Stat*, 2019; 73, 3, 307-309.
26. Bompa T, Theory and methodology of training, Kendall/Hunt Publishing Company, Toronto, 1994
27. Chatard, JC., Lavoie, JM., Lacour, JR. Swimming skill cannot be interpreted directly from the energy cost of swimming. In: D. Maclaren, T. Reilly, & A. Lees (Eds.), *Biomechanics and Medicine in Swimming* VI. London: E & FN Spon. 1992; pp. 173-180.
28. Hollander, P, de Groot, G, van Ingen Schenau, GJ, Kahman, R, Toussaint, HM, Contribution of the legs in front crawl swimming, in *Swimming Science.* Ungerechts, VBE, Reischle, K, Wilke, K Eds., , Human Kinetics, Champaign, Ill, USA, 1988; pp. 39-43
29. Shahbazi-Moghadam, M. Determination of arms and legs contribution to propulsion and percentage of coordination in breaststroke swim. In ISBS-Conference Proceedings Archive, 2008; 1,1.
30. Toussaint, HM, Beek PJ, Biomechanics of competitive front crawl swimming, *Sports Med*, 1992;13. 1, 8–24.
31. Duche, P, Falgairette, G, Bedu, M, Lac, G, Robert, A, & Coudert, J. Analysis of performance of prepubertal swimmers assessed from anthropometric and bio-energetic characteristics. *Eur J Appl Physiol Occup Physiol*.1993; 66: 467–471.
32. Mclean, SP., Hinrichs, RN. Sex differences in the centre of buoyancy location of competitive swimmers. *J Sports Sci.* 1998;*16*, 4, 373-383.
33. Sprague, HA. Relationship of certain physical measurements to swimming speed. *Res Q.* 1976; 47: 810–814,
34. Heinlein SA, Cosgarea AJ, Biomechanical considerations in the competitive swimmer's shoulder. *Sports Health*. 2010; 2: 519–525.

Table 1. Bayesian regression model predicting log-transformed swimming speeds across four swimming strokes with predictors that have the highest weight of evidence given the data as determined by Bayes Factor

	Ln (100 m speed [m.s^{-1}])		
Predictors	*Estimates*	*CI (95%)*	*Probability of direction (%)*
Intercept	-4.29	-4.98 – -3.56	100.00
Breaststroke	-0.11	-0.14 – -0.08	100.00
Butterfly	0.01	-0.02 – 0.03	63.80
Front crawl	0.15	0.12 – 0.17	100.00
Sex (Male)	0.03	0.02 – 0.05	100.00
Cubic (age) 1	0.58	0.33 – 0.82	100.00
Cubic (age) 2	-0.35	-0.50 – -0.19	100.00
Cubic (age) 3	0.26	0.14 – 0.38	100.00
Ln (Body Fat [%])	-0.09	-0.12 – -0.06	100.00
Ln (Bi-acromial breadth [cm])	0.42	0.25 – 0.59	100.00
Ln (Bi-iliac breadth [cm])	0.07	-0.02 – 0.17	96.45
Ln (Arm Span [cm])	0.33	0.18 – 0.47	100.00
Ln (Seated height [cm])	0.30	0.16 – 0.45	99.95
Observations	363		
R^2 Bayes	0.856		

The reference category in the model is backstroke and other stroke coefficients are compared to this

Table 2. Bayesian regression models of log-transformed swimming speeds for the four main swimming strokes with predictors that have the highest weight of evidence given the data as determined by Bayes Factor and LOOIC.

Quantile	0.1		0.5		0.9	
	Estimates	CI (95%)	Estimates	CI (95%)	Estimates	CI (95%)
Front Crawl Predictors						
Intercept	-4.42	-5.94 – -3.35	-4.82	-6.04 – -3.34	-3.64	-4.69 – -2.25
Ln (body fat [%})	-0.11	-0.14 – -0.07	-0.08	-0.13 – -0.04	-0.12	-0.19 – -0.06
Ln (Bi-acromial Breadth [cm])	0.59	0.32 – 0.90	0.55	0.34 – 0.80	0.77	0.47 – 0.96
Ln (Height [cm])	0.57	0.30 – 0.87	0.67	0.32 – 0.95	0.32	-0.03 – 0.64
R^2 Bayes	***0.8***		***0.785***		***0.793***	
Breaststroke Predictors						
Intercept	-2.29	-3.67 – -1.11	-1.42	-2.73 – -0.24	-2.71	-3.63 – -1.55
Quadratic (Age) 1	0.39	0.15 – 0.74	0.55	0.35 – 0.71	0.53	0.37 – 0.72
Quadratic (Age) 2	-0.06	-0.21 – 0.09	-0.13	-0.25 – -0.01	-0.1	-0.21 – 0.01
Ln (body fat [%])	-0.08	-0.17 – -0.00	-0.15	-0.22 – -0.08	-0.12	-0.18 – -0.06
Ln (Bi-acromial Breadth [cm])	0.25	-0.20 – 0.79	0.31	-0.00 – 0.66	0.62	0.35 – 0.82
Ln (Bi-iliac Breadth [cm])	0.47	0.11 – 0.78	0.23	0.04 – 0.44	0.27	0.09 – 0.44
R^2 Bayes	***0.645***		***0.721***		***0.779***	
Backstroke Predictors						
Intercept	-4.07	-4.90 – -3.27	-4.26	-5.20 – -3.32	-4.3	-5.28 – -3.35
Cubic(age) 1	1.6	1.31 – 1.91	1.27	0.97 – 1.56	0.94	0.60 – 1.27
Cubic(age) 2	-0.61	-0.88 – -0.37	-0.65	-0.83 – -0.47	-0.42	-0.62 – -0.21
Cubic (age) 3	0.06	-0.18 – 0.29	0.32	0.15 – 0.48	0.23	0.06 – 0.38
Ln (Body Fat [%])	-0.13	-0.17 – -0.09	-0.14	-0.18 – -0.10	-0.1	-0.14 – -0.05
Ln (Seated Height [cm])	0.15	0.01 – 0.31	0.22	0.05 – 0.40	0.49	0.25 – 0.62
Ln (Wrist Girth [cm])	0.04	-0.09 – 0.21	0.27	0.12 – 0.37	0.18	0.12 – 0.29
Ln (Calf Girth [cm])	0.21	0.05 – 0.41	0.24	0.11 – 0.39	0.20	0.04 – 0.38
Ln (Bi-iliac Breadth [cm)	0.18	0.05 – 0.32	0.22	0.11 – 0.32	0.28	0.11 – 0.42
ln (Arm Span [cm])	0.48	0.25 – 0.71	0.33	0.10 – 0.52	0.15	-0.08 – 0.37
R^2 Bayes	***0.781***		***0.786***		***0.728***	
Butterfly Predictors						
Intercept	-2.7	-3.71 – -1.90	-3.21	-4.24 – -2.19	-1.78	-2.76 – -0.86
Quadratic (Age) 1	0.87	0.61 – 1.07	0.47	0.25 – 0.72	0.69	0.45 – 0.92
Quadratic (Age) 2	-0.28	-0.48 – -0.05	-0.12	-0.32 – 0.05	-0.16	-0.29 – -0.01
Ln (body fat [%})	-0.16	-0.20 – -0.12	-0.18	-0.24 – -0.11	-0.15	-0.19 – -0.10
Ln (Calf Girth [cm])	0.60	0.27 – 0.92	0.82	0.48 – 1.10	0.45	0.25 – 0.67
Ln (Ankle Girth [cm])	-0.56	-0.78 – -0.23	-0.41	-0.68 – -0.13	-0.20	-0.46 – 0.03
Ln (Bi-acromial Breadth [cm])	0.62	0.37 – 0.93	0.51	0.19 – 0.84	0.18	-0.08 – 0.48
Ln (Bi-iliac Breadth [cm)	0.16	-0.03 – 0.34	0.14	-0.02 – 0.31	0.27	0.12 – 0.45
R^2 Bayes	***0.727***		***0.691***		***0.624***	

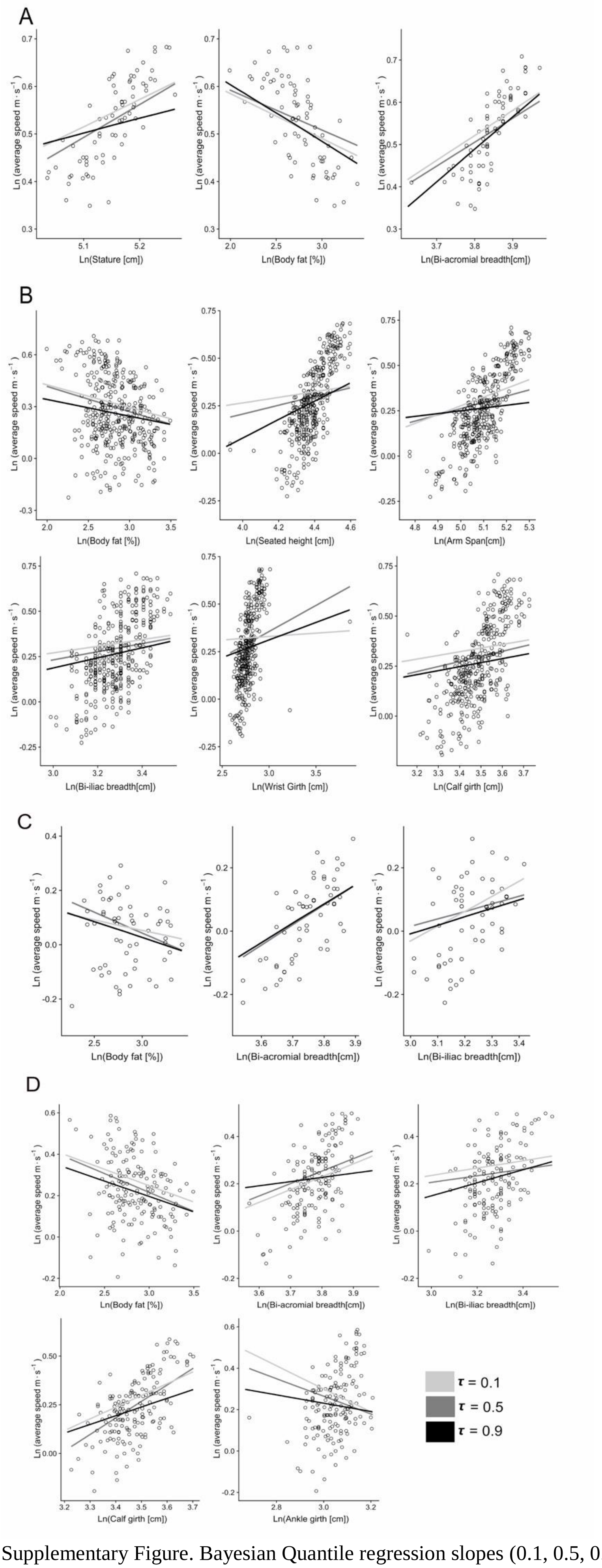
Table 4. Selected predictions for the best performers (0.9 quantile) from the models for each swimming stroke for 100-m performance for a male 15-year-old swimmer

Front Crawl †		Speed ($m.s^{-1}$)	Time for 100-m (sec)
Bi-acromial breadth (cm)	Height (cm)		
55	195	1.99	50.34
54	195	1.96	51.11
53	195	1.93	51.84
55	190	1.96	50.91
54	190	1.94	51.61
53	190	1.91	52.26
Breaststroke †		Speed ($m.s^{-1}$)	Time for 100-m (sec)
Bi-acromial breadth (cm)	Bi-iliac breadth (cm)		
55	35	1.43	70.17
55	33	1.40	71.30
55	31	1.38	72.61
50	35	1.35	74.17
50	33	1.33	75.33
50	31	1.30	76.68
Backstroke †		Speed ($m.s^{-1}$)	Time for 100-m (sec)
Seated height (cm)	Bi-iliac breadth (cm)		
100	35	1.58	63.20
100	33	1.56	64.16
100	30	1.53	65.53
98	35	1.57	63.80
98	33	1.55	64.67
98	30	1.50	64.67
Butterfly †		Speed ($m.s^{-1}$)	Time for 100-m (sec)
Calf Girth (cm)	Ankle Girth (cm)		
42	14	1.60	62.68
40	14	1.56	64.11
38	14	1.53	65.47
42	15	1.57	63.58
40	17	1.50	66.81
38	19	1.43	69.80

† *Variables in the model but not included in the table were held at their empirical mean*

Supplementary Table . Descriptive statistics (means ± SD) of swimming speed, demographic and somatic measurements by sex and stroke

Variables	Male								Female							
Strokes	Breaststroke	SD	Butterfly	SD	Backstroke	SD	Front Crawl	SD	Breaststroke	SD	Butterfly	SD	Backstroke	SD	Front Crawl	SD
N	39		103		30		30		20		64		33		44	
100 m time (s)	97.7	13.5	79.1	13.2	77.2	8.8	54.8	3.3	95.4	9.5	81.8	11.2	79.5	5.0	62.0	4.5
Speed ($m.s^{-1}$)	1.04	0.14	1.30	0.22	1.31	0.17	1.83	0.11	1.06	0.10	1.24	0.16	1.26	0.08	1.62	0.12
Age (yrs)	11.5	1.3	13.1	2.8	14.0	0.6	19.0	3.8	12.1	1.0	13.6	2.6	13.0	1.2	15.9	2.7
Body Mass (kg)	41.5	9.5	50.0	14.2	48.7	12.3	73.2	9.7	46.0	8.6	49.3	9.2	49.1	7.1	62.5	9.2
Height (cm)	149.9	10.4	158.3	12.7	157.2	11.6	177.8	6.5	155.9	8.0	157.9	9.0	158.8	7.4	168.9	9.0
Body Fat (%)	16.8	5.5	16.6	5.2	17.6	5.7	12.9	2.8	19.0	4.3	18.6	3.7	18.2	3.9	18.0	4.4
Sitting height (cm)	74.4	5.8	78.6	7.1	77.8	6.8	90.7	4.3	76.4	7.9	79.0	6.6	79.7	3.9	85.2	5.0
Upper limb length (cm)	69.3	5.3	73.3	6.2	72.6	5.6	82.6	3.5	72.0	4.4	73.1	4.5	74.1	4.4	78.6	5.0
Upper arm length (cm)	29.0	2.2	30.8	2.9	30.5	2.5	34.9	2.1	30.4	2.0	31.0	2.3	31.2	2.2	33.3	2.6
Lower arm length (cm)	23.2	1.9	24.3	2.1	23.9	1.6	27.2	1.8	23.7	1.7	24.2	1.7	24.3	1.7	25.7	1.9
Hand length (cm)	18.0	1.6	19.0	1.6	18.9	1.3	21.1	1.1	18.7	1.0	18.9	1.0	19.0	0.9	19.9	1.2
Lower limb length (cm)	81.6	6.1	85.9	6.6	85.5	5.6	93.4	4.2	85.0	5.0	86.4	5.7	86.9	5.0	92.1	5.2
Thigh length (cm)	38.9	5.8	41.4	3.1	41.6	3.1	44.3	2.3	42.0	2.2	42.8	3.4	42.9	3.0	45.5	3.9
Leg length (cm)	42.0	3.4	44.2	3.5	43.7	2.8	48.2	2.7	43.4	2.5	43.7	2.7	44.1	2.5	46.1	2.9
Foot length (cm)	25.1	2.2	26.2	1.8	26.1	1.6	27.8	1.3	25.5	1.1	25.2	1.1	25.3	1.1	26.3	1.5
Arm relaxed Girth (cm)	23.0	3.0	24.8	3.8	24.4	3.2	30.7	2.9	23.8	2.3	24.6	2.2	24.1	2.0	28.1	2.0
Forearm Girth (cm)	21.1	1.9	22.7	3.0	22.2	2.4	27.5	1.9	21.5	1.5	21.9	1.6	21.8	1.5	24.3	2.0
Wrist Girth (cm)	14.7	1.1	15.5	1.4	15.4	1.3	17.4	0.9	15.4	2.3	15.0	0.8	15.2	0.8	16.8	4.9
Thigh Girth (cm)	43.9	4.5	46.8	5.4	46.5	4.9	52.9	5.0	46.1	5.6	47.7	4.4	47.3	3.8	51.8	4.3
Calf Girth (cm)	30.3	3.1	32.3	3.4	32.4	3.1	36.7	2.0	30.7	3.0	31.9	2.6	32.4	2.3	34.9	2.8
Ankle Girth (cm)	20.5	1.8	21.2	1.8	21.2	2.1	23.1	1.3	20.5	1.5	20.7	1.2	20.9	1.1	22.4	1.4
Bi-acromial breadth (cm)	41.5	3.9	43.8	3.5	43.4	2.6	48.2	2.4	43.3	2.3	43.7	2.1	43.9	2.0	45.3	2.4
Bi-iliac breadth (cm)	24.4	2.4	26.2	2.7	25.7	2.3	29.0	2.2	25.7	2.4	26.5	2.3	26.4	2.2	28.6	2.3
Arm span (cm)	150.4	13.5	161.1	15.6	160.6	13.1	184.5	9.8	158.3	9.7	160.6	10.6	160.7	10.2	173.4	11.6



Supplementary Figure. Bayesian Quantile regression slopes (0.1, 0.5, 0.9) for the four main swimming s